A simple model of global cascades on random networks

Duncan J. Watts*

Department of Sociology, Columbia University New York, NY 10027

Communicated by Murray Gell-Mann, Santa Fe Institute, Santa Fe, NM, February 14, 2002 (received for review May 29, 2001)

The origin of large but rare cascades that are triggered by small initial shocks is a phenomenon that manifests itself as diversely as cultural fads, collective action, the diffusion of norms and innovations, and cascading failures in infrastructure and organizational networks. This paper presents a possible explanation of this phenomenon in terms of a sparse, random network of interacting agents whose decisions are determined by the actions of their neighbors according to a simple threshold rule. Two regimes are identified in which the network is susceptible to very large cascades—herein called global cascades—that occur very rarely. When cascade propagation is limited by the connectivity of the network, a power law distribution of cascade sizes is observed, analogous to the cluster size distribution in standard percolation theory and avalanches in self-organized criticality. But when the network is highly connected, cascade propagation is limited instead by the local stability of the nodes themselves, and the size distribution of cascades is bimodal, implying a more extreme kind of instability that is correspondingly harder to anticipate. In the first regime, where the distribution of network neighbors is highly skewed, it is found that the most connected nodes are far more likely than average nodes to trigger cascades, but not in the second regime. Finally, it is shown that heterogeneity plays an ambiguous role in determining a system's stability: increasingly heterogeneous thresholds make the system more vulnerable to global cascades; but an increasingly heterogeneous degree distribution makes it less vulnerable.

How is it that small initial shocks can cascade to affect or disrupt large systems that have proven stable with respect to similar disturbances in the past? Why do some books, movies, and albums emerge out of obscurity, and with small marketing budgets, to become popular hits (1), when many *a priori* indistinguishable efforts fail to rise above the noise? Why does the stock market exhibit occasional large fluctuations that cannot be traced to the arrival of any correspondingly significant piece of information (2)? How do large, grassroots social movements start in the absence of centralized control or public communication (3)?

These phenomena are all examples of what economists call *information cascades* (ref. 4; but which are herein called simply *cascades*), during which individuals in a population exhibit herd-like behavior because they are making decisions based on the actions of other individuals rather than relying on their own information about the problem. Although they are generated by quite different mechanisms, cascades in social and economic systems (3–6) are similar to cascading failures in physical infrastructure networks (7, 8) and complex organizations (9) in that initial failures increase the likelihood of subsequent failures, leading to eventual outcomes that, like the August 10, 1996 cascading failure in the western United States power transmission grid (8), are extremely difficult to predict, even when the properties of the individual components are well understood. Not as newsworthy, but just as important as the cascades themselves, is that the very same systems routinely display great stability in the presence of continual small failures and shocks that are at least as large as the shocks that ultimately generate a cascade. Cascades can therefore be regarded as a specific manifestation of the *robust yet fragile* nature of many complex systems (10): a system may appear stable for long periods of time and withstand many external shocks (robust), then suddenly and apparently inexplicably exhibit a large cascade (fragile).

Although the social, economic, and physical mechanisms responsible for the occurrence of cascades are complex and may vary widely across systems and even between particular cascades in the same system, it is proposed in this paper that some generic features of cascades can be explained in terms of the connectivity of the network by which influence is transmitted between individuals. Specifically, this paper addresses the set of qualitative observations that (*i*) global (i.e., very large) cascades can be triggered by exogenous events (shocks) that are very small relative to the system size, and (*ii*) global cascades occur rarely relative to the number of shocks that the system receives, and may be triggered by shocks that are *a priori* indistinguishable from shocks that do not.

Model Motivation: Binary Decisions with Externalities

This model is motivated by considering a population of individuals each of whom must decide between two alternative actions, and whose decisions depend explicitly on the actions of other members of the population. In social and economic systems, decision makers often pay attention to each other either because they have limited information about the problem itself or limited ability to process even the information that is available (6). When deciding which movie (11) or restaurant (12) to visit, we often have little information with which to evaluate the alternatives, so frequently we rely on the recommendation of friends, or simply pick the movie or restaurant to which most people are going. Even when we have access to plentiful information, such as when we evaluate new technologies, risky financial assets, or job candidates, we often lack the ability to make sense of it; hence, again we rely on the advice of trusted friends, colleagues, or advisors. In other decision making scenarios, such as in collective action problems (3) or social dilemmas (13), an individual's payoff is an explicit function of the actions of others. And in other problems still, involving say the diffusion of a new technology (14), the utility of a single additional unit—a fax machine for example—may depend on the number of units that have already been sold. In all these problems, therefore, regardless of the details, individual decision makers have an incentive to pay attention to the decisions of others.

In economic terms, this entire class of problems is known generically as *binary decisions with externalities* (6). As simplistic as it appears, a binary decision framework is relevant to surprisingly complex problems. To take an extreme example, the creation of a political coalition or an international treaty is unquestionably a complex, multifaceted process with many potential outcomes. But once the coalition exists or the treaty has been drafted, the decision of whether or not to join is essential a binary one. Similar reasoning applies to a firm's choice between two technologies, or an individual's choice between two neighborhood restaurants—the factors involved in the decision may be many, but the decision itself can be regarded as binary.

Both the detailed mechanisms involved in binary decision problems, and also the origins of the externalities can vary widely across specific problems. Nevertheless, in many applications that have been examined in the economics and sociology literature—for

*E-mail: djw24@columbia.edu.

example, fads (1, 4, 5), riots (15), crime (16), competing technologies (14), and the spread of innovations (17, 18), conventions (6), and cooperation (13)—the decision itself can be considered a function solely of the *relative* number of other agents who are observed to choose one alternative over the other (6). Because many decisions are inherently costly, requiring commitment of time or resources, the relevant decision function frequently exhibits a strong *threshold* nature: agents display inertia in switching states, but once their personal threshold has been reached, the action of even a single neighbor can tip them from one state to another.

Model Specification

A particularly simple binary decision rule with externalities that captures the essential features outlined above is the following: An individual agent observes the current states (either 0 or 1) of k other agents, which we call its *neighbors*, and adopts state 1 if at least a threshold fraction ϕ of its k neighbors are in state 1, else it adopts state 0.

To account for variations in knowledge, preferences, and observational capabilities across the population of decision-making agents, both individual thresholds and also the number of neighbors k are allowed to be heterogeneous. First, each agent is assigned a threshold ϕ drawn at random from a distribution $f(\phi)$ that is defined on the unit interval and normalized such that $\int_0^1 f(\phi)d\phi = 1$, but which is otherwise arbitrary. Next, we construct a network of n agents, in which each agent is connected to k neighbors with probability p_k and the average number of neighbors is $\langle k \rangle = z$. Although we shall continue to speak of an agent's *neighbors*, we should think of them simply as the set of incoming signals that are relevant to the problem at hand. More formally, we say that agents are represented by *vertices* (or *nodes*) in a graph; neighboring vertices are joined by *edges*; p_k is the *degree distribution* of the graph; and z is the *average degree* (in physics, z is usually called the *coordination number*). To model the dynamics of cascades, the population is initially all-off (state 0) and is perturbed at time $t = 0$ by a small fraction $\Phi_0 \ll 1$ of vertices that are switched on (state 1). The population then evolves at successive time steps with all vertices updating their states in random, asynchronous order according to the threshold rule above. Once a vertex has switched on, it remains on (*active*) for the duration of the dynamics.

In the social science literature, decision rules of this kind are usually derived either from the payoff structure of noncooperative games such as the prisoner's dilemma (3, 6), or from stochastic sampling procedures (18). But when regarded more generally as a change of state—not just a decision—the model belongs to a larger class of contagion problems that includes models of failures in engineered systems such as power transmission networks (8) or the internet (19, 20), epidemiological (21) and percolation (22, 23) models of disease spreading, and a multiplicity of cellular-automata models including random-field Ising models (24), bootstrap percolation (25, 26), majority voting (27, 28), spreading activation (29), and self-organized criticality (8, 29).

The model, however, differs from these other contagion models in some important respects. (*i*) Unlike epidemiological models, where contagion events between pairs of individuals are *independent*, the threshold rule effectively introduces *local dependencies*; that is, the effect that a single infected neighbor will have on a given node depends critically on the states of the node's other neighbors. (*ii*) Unlike bootstrap percolation, and self-organized criticality models (which also exhibit local dependencies), the threshold is not expressed in terms of the absolute number of a node's neighbors choosing a given alternative, but the corresponding fraction of the neighborhood. This is a natural condition to impose for decision making problems, because the more signals a decision maker receives, the less significant any one signal becomes. (*iii*) Unlike random-field Ising and majority vote models, which are typically modeled on regular lattices, here we are concerned with heterogeneous networks; that is, networks in which individuals have different numbers of neighbors. All these features—*local dependencies*, *fractional thresholds*, and *heterogeneity*—are essential to the dynamics of cascades. Furthermore, although they are clearly related by the threshold condition, network heterogeneity and threshold heterogeneity turn out not to be equivalent, and therefore need to be considered separately.

Exact Solution on an Arbitrary Random Graph

The main objective of this paper is to explore how the vulnerability of interconnected systems to global cascades depends on the network of interpersonal influences governing the information that individuals have about the world, and therefore their decisions. Because building relationships and gathering information are both costly exercises, interaction and influence networks tend to be very sparse (17)—a characteristic that appears to be true of real networks in general (30)—so we consider only the properties of networks with $z \ll n$. In the absence of any known geometry for the problem, a natural first choice for a sparse interaction network is an undirected random graph (31), with n vertices and specified degree distribution p_k. Although random graphs are not considered to be highly realistic models of most real-world networks (30), they are often used as first approximations (19, 20, 32) because of their relative tractability, and this tradition is followed here. Our approach concentrates on two quantities: (*i*) the probability that a *global cascade* will be triggered by a single node (or small seed of nodes), where we define a global cascade formally as cascade that occupies a finite fraction of an infinite network; and (*ii*) the expected size of a global cascade once it is triggered. When describing our results, the term *cascade* therefore refers to an event of any size triggered by an initial seed, whereas *global cascade* is reserved for *sufficiently large* cascades (in practice, this means more than a fixed fraction of large, but finite network).

In any sufficiently large random graph with $z < c \ln n$ (where c is some constant) and $\Phi_0 \ll 1$ (i.e., sparsely connected with a small initial seed), we can assume that the local neighborhood of a small seed will not contain any short cycles; hence, no vertex neighboring the initial seed will be adjacent to more than one seed member. This approximation becomes exact in the case of an infinite network, with finite z, or a seed consisting of a single vertex. Under this condition, the only way in which the seed can grow is if at least one of its immediate neighbors has a threshold such that $\phi \le 1/k$, or equivalently has degree $k \le K = \lfloor 1/\phi \rfloor$. We call vertices that are unstable in this one-step sense, *vulnerable*, and those that are not, *stable*, noting that the distinction only applies when the seed in question is *small* (numerical simulations suggest that seeds that are three orders of magnitude less than the system size are sufficiently small). The case of large seeds will be discussed later.

Although the vulnerability condition is quite general, for concreteness we use the language of the diffusion of innovations (17), in which the initial seed plays the role of the *innovators*, and vulnerable vertices correspond to *early adopters*. Unless the innovators are connected to a community of early adopters, no cascade is possible. In fact, as we show below, the success or failure of an innovation may depend less on the number and characteristics of the innovators themselves than on the structure of the community of early adopters. Clearly, the more early adopters exist in the network, the more likely it is that an innovation will spread. But the extent of its growth—and hence the susceptibility of the network as a whole—depends not only on the number of early adopters, but on how connected they are to one another, and also to the much larger community consisting of the *early and late majority*, who do not tend to respond to the innovators directly, but who can be influenced indirectly if exposed to multiple early adopters. In the context of this model, we conjecture that the required condition for a global cascade is that the subnetwork of vulnerable vertices must *percolate* (22) throughout the network as a whole, which is to say that the largest, connected vulnerable cluster must occupy a finite fraction of an infinite network. Regardless of how connected the network as a whole might be, the claim here is that only if the largest *vulnerable* cluster percolates are global cascades possible.

This condition, which we call the *cascade condition* (see Eq. **5** below), has the considerable advantage of reducing a complex dynamics problem to a static, percolation problem that can be solved using a generating function approach. A similar technique has been used elsewhere (20, 32) to study the connectivity properties of random graphs; here the basic approach is modified (described in detail in ref. 32) to focus on vulnerable vertices. By construction, every vertex has degree k with probability p_k, and by the vulnerability condition above, a vertex with degree k is vulnerable with probability $\rho_k = P[\phi \leq 1/k]$. Hence, the probability of a vertex u having degree k and being vulnerable is $\rho_k p_k$, and the corresponding generating function of vulnerable vertex degree is:

$$G_0(x) = \sum_k \rho_k p_k x^k, \quad \textbf{[1a]}$$

$$\text{where } \rho_k = \begin{cases} 1 & k = 0 \\ F(1/k) & k > 0 \end{cases} \quad \textbf{[1b]}$$

and $F(\phi) = \int_0^{\phi} f(\varphi)d\varphi$. By incorporating all of the information contained in the degree distribution and the threshold distribution, $G_0(x)$ generates all of the moments of the degree distribution solely of *vulnerable* vertices, where the relevant moments can be extracted by evaluating the derivatives of $G_0(x)$ at $x = 1$. For the purposes of this paper, the two most important quantities are (*i*) the vulnerable fraction of the population $P_v = G_0(1)$, and (*ii*) the average degree of vulnerable vertices $z_v = G_0'(1)$. Because we are interested in the propagation of cascades from one vertex to another, we also require the degree distribution of a vulnerable vertex v that is a random neighbor of our initially chosen vertex u. The larger the degree of v, the more likely it is to be a neighbor of u; hence, the probability of choosing v is proportional to kp_k, and the correctly normalized generating function $G_1(x)$ corresponding to a neighbor of u is:

$$G_1(x) = \frac{\sum_k k \rho_k p_k x^{k-1}}{\sum_k k p_k} = \frac{G_0'(x)}{z}. \quad \textbf{[2]}$$

To calculate the properties of clusters of vulnerable vertices (the community structure of the early adopters), we introduce the analogous generating functions

$$H_0(x) = \sum_n q_n x^n \text{ and } H_1(x) = \sum_n r_n x^n,$$

where q_n is the probability that a randomly chosen vertex will belong to a vulnerable cluster of size n, and r_n is the corresponding probability for a neighbor of an initially chosen vertex. Any *finite* cluster of size n that we arrive at by following a random edge can be regarded as composed of smaller such clusters, whose cumulative sizes must sum to n. Because a sufficiently large random graph below percolation can be regarded as a pure branching structure, we can therefore ignore the possibility that the subclusters will be connected in cycles, so each subcluster can be treated independently of the others. (The presence of an infinite cluster above percolation will be dealt with below.) Hence, the probability of a finite cluster of size n is simply the product of the probabilities of its (also finite) subclusters. It follows from the properties of generating functions (20, 32) that $H_1(x)$ satisfies the following self-consistency equation:

$$H_1(x) = [1 - G_1(1)] + xG_1(H_1(x)), \quad \textbf{[3a]}$$

from which $H_0(x)$ can be computed according to

$$H_0(x) = [1 - G_0(1)] + xG_0(H_1(x)). \quad \textbf{[3b]}$$

where the first term in both Eqs. **3a** and **3b** corresponds to the probability that the vertex chosen is not vulnerable, and the second term accounts for the size distribution of vulnerable clusters attached to a vertex that is, itself, vulnerable. $H_0(x)$ therefore generates all moments of the distribution of vulnerable cluster sizes, the most important of which, for our current purpose, is the *average vulnerable cluster size* $\langle n \rangle = H_0'(1)$, because this is the quantity that diverges at percolation. Substituting the expressions for $H_0(x)$ and $H_1(x)$ above, we find that

$$\langle n \rangle = G_0(1) + (G_0'(1))^2/(z - G_0''(1)) = P + z_v^2/(z - G_0''(1)), \quad \textbf{[4]}$$

which diverges when

$$G_0''(1) = \sum_k k(k-1)\rho_k p_k = z. \quad \textbf{[5]}$$

Eq. **5**—the *cascade condition*—is interpreted as follows: When $G_0''(1) < z$, all vulnerable clusters in the network are small; hence, the early adopters are isolated from each other and will be unable to generate the momentum necessary for a cascade to become global. But when $G_0''(1) > z$, the typical size of vulnerable clusters is infinite, implying the presence of a percolating vulnerable cluster, in which case random initial shocks should trigger global cascades with finite probability. Because Eq. **5** marks the transition between these two regimes, or *phases*, at which the average cluster size diverges and global cascades first commence, it is called a *phase transition* (31–33). The conditions necessary to generate global cascades can, in other words, be determined by locating the position and nature of the relevant phase transition. Note, however, that the $k(k-1)$ term in Eq. **5** is monotonically increasing in k, but ρ_k is monotonically decreasing. Thus we would expect that Eq. **5** will have either two solutions (resulting in two phase transitions), or none at all, in contrast with the usual percolation model, which exhibits a single phase transition in z for all finite values of the occupation probability. Furthermore, in the case where we have two solutions, we should observe a continuous interval in z, inside which cascades occur.

Results and Discussion

Although the cascade condition (Eq. **5**) applies to random graphs with arbitrary degree distributions p_k and threshold distributions $f(\phi)$ (expressed through the weighting function ρ_k), we can illustrate its main features for the special case of a uniform random graph (in which any pair of vertices is connected with probability $p = z/n$), and where all vertices have the same threshold ϕ; that is, $f(\phi) = \delta(\phi - \phi_*)$. A characteristic of uniform random graphs is that $p_k = e^{-z}z^k/k!$ (the Poisson distribution), in which case Eq. **5** reduces to $zQ(K_* - 1, z) = 1$, where $K_* = \lfloor 1/\phi_* \rfloor$ and $Q(a, x)$ is the incomplete gamma function. Fig. 1 expresses the cascade condition graphically as a boundary in the (ϕ_*, z) phase diagram (dashed line) and compares it to the region (outlined by solid circles) in which cascades are observed over 1,000 realizations of the dynamics (each realization consists of a randomly constructed network of 10,000 vertices, in which a single vertex is switched on at $t = 0$). Because the simulated system is finite, the predicted and actual boundaries of the cascade window do not agree perfectly, but they are very similar. In particular, as predicted above, both display a lower and an upper boundary as a function of the average degree z, at which the characteristic time scale of the dynamics diverges (see Fig. 2*a*).

To understand the nature of the phase transitions that define the boundaries of the cascade window, we solve exactly for the fractional size S_v of the vulnerable cluster inside the cascade window. Because the generation function approach requires the largest vulnerable cluster to be a pure branching structure, and because the vulnerable cluster will, in general, contain cycles above percolation, Eq. **4** only applies below percolation, which is why Eq. **5** can only specify the boundary of the cascade window. However, we can still solve for S_v above the phase transition, as well as below it, by evaluating $H_0(1)$ exclusively over the set of finite clusters; that is, by

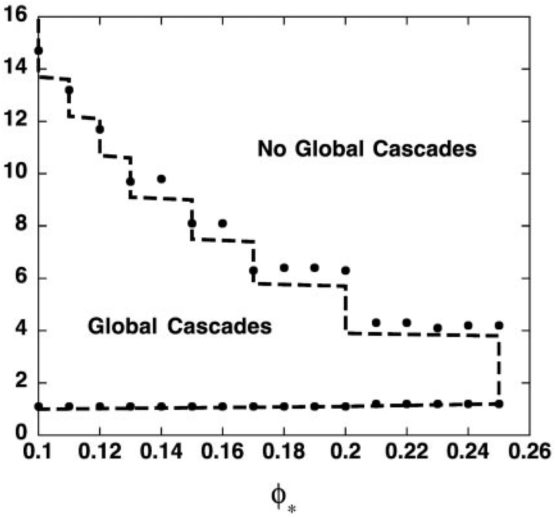

Fig. 1. Cascade windows for the threshold model. The dashed line encloses the region of the (ϕ_*, z) plane in which the cascade condition (Eq. **5**) is satisfied for a uniform random graph with a homogenous threshold distribution $f(\phi) = \delta(\phi - \phi_*)$. The solid circles outline the region in which global cascades occur for the same parameter settings in the full dynamical model for $n = 10{,}000$ (averaged over 100 random single-node perturbations).

explicitly excluding the percolating cluster (when it exists) from the sum $\Sigma_n q_n x^n$. Using Eq. **3b**, it follows that $S_v = 1 - H_0(1) = P - G_0(H_1(1))$, where $H_1(1)$ satisfies Eq. **3a**. Outside the cascade window, the only solution to Eq. **3a** is $H_1(1) = 1$, which yields $S_v = 0$ (and therefore no cascades) as expected. But inside the cascade window, where the percolating vulnerable cluster exists, Eq. **3a** has an additional solution that corresponds to a non-zero value of S_v. In the special case of a uniform random graph with homogeneous thresholds, we obtain $S_v = Q(K_* + 1, z) - e^{z(H_1 - 1)}Q(K_* + 1, zH_1)$, in which H_1 satisfies $H_1 = 1 - Q(K_*, z) + e^{z(H_1 - 1)}Q(K_*, zH_1)$. We contrast this expression with that for the size of the entire connected component of the graph, $S = 1 - e^{-zS}$ (32), which is equivalent to allowing $K_* \to \infty$ (or $\phi_* \to 0$). In Fig. 2*b* we show the exact solutions for both S_v (long-dashed line) and S (solid line) for the case of $\phi_* = 0.18$, and compare these quantities with the frequency and size of global cascades observed in the full dynamical simulation of 10,000 nodes averaged over 1,000 random realizations of the network and the initial condition. (The corresponding numerical values for S_v and S are indistinguishable from the analytical curves, except near the upper boundary of the window.)

The frequency of global cascades (open circles)—that is, cascades that are "successful"—is obviously related to the size of the vulnerable component: the larger is S_v, the more likely a randomly chosen initial site is to be a part of it. In particular, if S_v does not percolate, then global cascades are impossible. Fig. 2*b* clearly supports this intuition, but it is equally clear that, within the cascade window, S_v seriously underestimates the likelihood of a global cascade. The reason is that, according to our original decision rule, an individual's choice of state depends only on the states of its neighbors; hence, even stable vertices, although they do not participate in the initial stages of a global cascade, can still trigger them as long as they are directly *adjacent* to the vulnerable cluster. The true likelihood of a global cascade is therefore determined by the size of what we call the *extended vulnerable cluster* S_e, consisting of the vulnerable cluster itself, and any stable vertices immediately adjacent to it. We have not solved for S_e exactly (although this may be possible), but it is relatively simple to determine numerically, and as the corresponding (dotted) curve in Fig. 2*b* demonstrates, the average value of S_e is an excellent approximation to the observed frequency of global cascades.

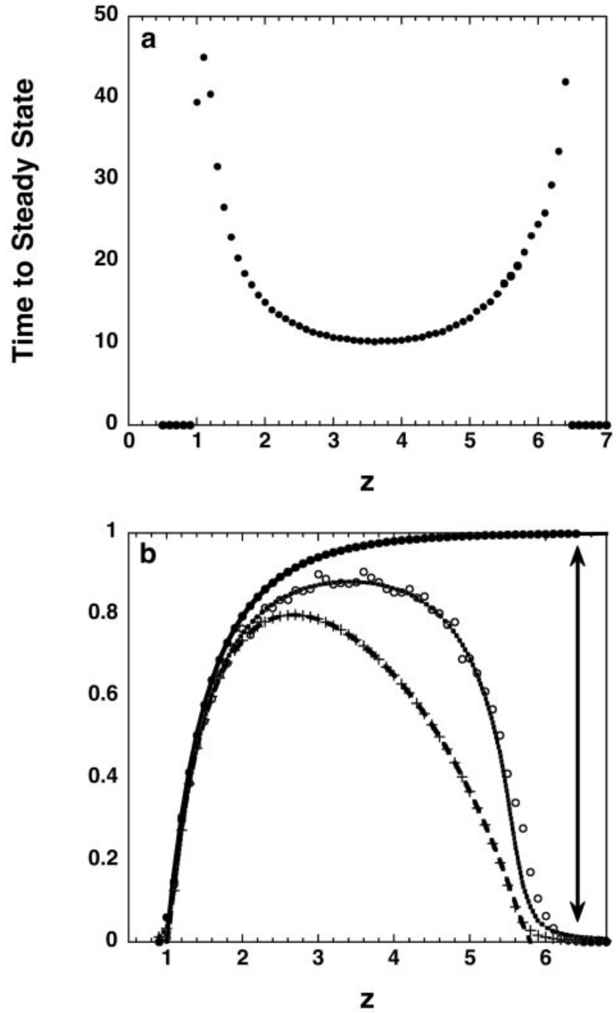

Fig. 2. Cross section of the cascade window from Fig. 1, at $\phi_* = 0.18$. (*a*) The average time required for a cascade to terminate diverges at both the lower and upper boundaries of the cascade window, indicating two phase transitions. (*b*) Comparison between connected components of the network and the properties of global cascades. The frequency of global cascades in the numerical model (open circles) is well approximated by the fractional size of the *extended vulnerable cluster* (short dashes). For comparison, the size of the *vulnerable cluster* is also shown, both the exact solution derived in the text (long dashes) and the average over 1,000 realizations of a random graph (crosses). The exact and numerical solutions agree everywhere except at the upper phase transition, where the finite size of the network ($n = 10{,}000$) affects the numerical results. Finally, the average size of global cascades is shown (solid circles) and compared with the exact solution for the largest connected component (solid line).

The average size of global cascades (solid circles) is clearly not governed either by the size of the vulnerable cluster S_v, or by S_e, but by S, the connectivity of the network as a whole. This is a surprising result, the reason for which is not entirely clear, but a plausible explanation is as follows. If a global cascade is triggered by an initially small seed striking the extended vulnerable cluster, it is guaranteed to occupy the entire vulnerable cluster, and therefore a finite fraction of even an infinite network. At this stage, the small-seed condition no longer holds, and so nodes that are still in the off state can now have multiple (early-adopting) neighbors in the on state. Hence, even individuals that were originally classified as stable (the early and late majority) can now be toppled, allowing the cascade to occupy not just the vulnerable component that allowed the cascade to spread initially, but the entire connected component of the graph. That the activation of a percolating

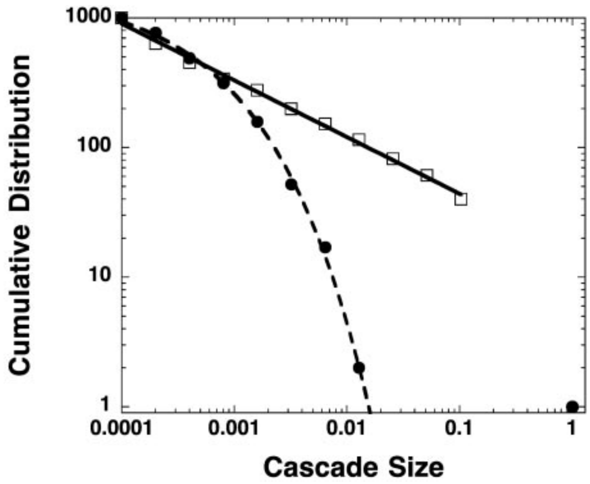

Fig. 3. Cumulative distributions of cascade sizes at the lower and upper critical points, for $n = 1{,}000$ and $z = 1.05$ (open squares) and $z = 6.14$ (solid circles), respectively. The straight line on the double logarithmic scale indicates that cascades at the lower critical point are power-law distributed, with slope 3/2 (the cumulative distribution has slope 1/2). By contrast, the distribution at the upper critical point is bimodal, with an exponential tail at small cascade size, and a second peak at the size of the entire system corresponding to a single global cascade. Above the upper boundary, the global cascade disappears and large cascades are always exponentially unlikely.

vulnerable cluster should always be sufficient to activate the entire connected component, even when the former is a very small fraction of the latter, is not an obvious result, but it appears to hold consistently, at least within the class of random graphs. Whether or not it turns out to hold for networks more general than random graphs is a matter of current investigation.

As Figs. 1 and 2 suggest, the onset of global cascades can occur in two distinct regimes—a low connectivity regime and a high connectivity regime—corresponding to the lower and upper phase transitions respectively. The nature of the phase transitions at the two boundaries is different, and this has important consequences for the apparent stability of the systems involved. As Fig. 3 (open squares) demonstrates, the cumulative distribution of cascades at the lower boundary of the cascade window follows a power law, analogous to the distribution of avalanches in models of self-organized criticality (29) or the cluster size distribution at criticality for standard percolation (22). In fact, the slope of the cascade size distribution is indistinguishable from the known critical exponent $\alpha = 3/2$ for the cluster size distribution of random graphs at percolation (32). This result is expected because, when $z \simeq 1$, most vertices satisfy the vulnerability condition, so the propagation of cascades is constrained principally by the connectivity of the network, which for random graphs is known to undergo a second-order phase transition at $z = 1$ (31).

The upper boundary, however, is different. Here, the propagation of cascades is limited not by the connectivity of the network, but by the local stability of the vertices. Most vertices in this regime have so many neighbors that they cannot be toppled by a single neighbor perturbation; hence, most initial shocks immediately encounter stable vertices. Most cascades therefore die out before spreading very far, giving the appearance that large cascades are exponentially unlikely. A percolating vulnerable cluster, however, still exists, so very rarely a cascade will be triggered in which case the high connectivity of the network ensures that it will be extremely large, typically much larger than cascades at the lower phase transition. The result is a distribution of cascade sizes that is bimodal rather than a power law (see Fig. 3, solid circles). As the upper phase transition is approached from below, global cascades become larger, but increasingly rare, until they disappear altogether, implying a discontinuous (i.e., first-order) phase transition in the size of successful cascades (see Fig. 2*b*, solid circles). The main consequence of the first-order phase transition is that just inside the boundary of the window, where global cascades occur very rarely (Fig. 3 shows only a single cascade occurring in 1,000 random trials), the system will in general be indistinguishable from one that is highly stable, exhibiting only tiny cascades for many initial shocks before generating a massive, global cascade in response to a shock that is *a priori* indistinguishable from any other.

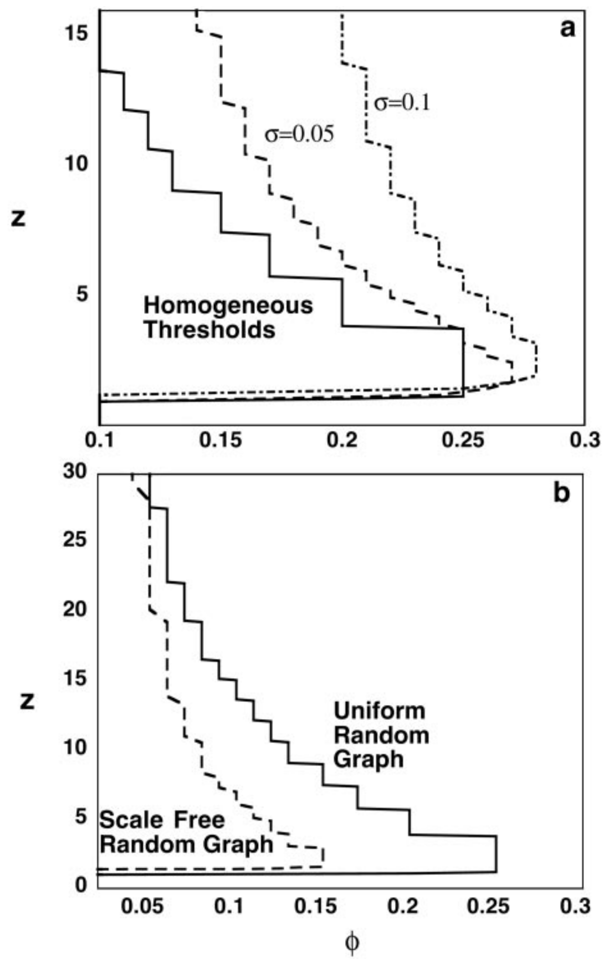

Fig. 4. Analytically derived cascade windows for heterogeneous networks. The solid lines are the same as Fig. 1. (*a*) The dashed lines represent cascade windows for uniform random graphs, but where the threshold distributions (ϕ) are normally distributed with mean ϕ and SD $\sigma = 0.05$ and $\sigma = 0.1$. (*b*) The dashed line represents the cascade window for a random graph with a degree distribution that is a power law with exponent τ and exponential cut-off κ_0, where τ has been fixed at $\tau = 2.5$ and κ_0 has been adjusted to generate graphs with variable z.

These qualitative results are quite general within the class of random networks, applying to arbitrary distributions both of thresholds $f(\phi)$ and degree p_k. Variations in either distribution, however, can affect the quantitative results—and thus the effective vulnerability of the system—considerably, as is demonstrated in Fig. 4 *a* and *b*. Fig. 4*a* shows the original cascade window for homogeneous thresholds (solid line) and also two windows (dashed lines) derived by the same generating function method, but corresponding to threshold distributions $f(\phi)$ that are normally distributed with mean ϕ_* and increasing standard deviation σ. Numerical results (not shown) correspond to the analytically derived windows. Clearly, increased heterogeneity of thresholds causes the system to be less stable, yielding cascades over a greater range of both ϕ_* and z. Fig. 4*b*, however, presents a different view of heterogeneity. Now the threshold distribution is held fixed, with all vertices exhibiting the same threshold, but the distribution of degree p_k is given by $p_k = Ck^{-\tau}e^{-k/\kappa}$ ($k > 0$), where C, τ, and κ are constants that can be adjusted such that we retain $\langle k \rangle = z$. This class of power-law random graphs has attracted much recent interest (19, 20, 32) as a model of many real networks, including the internet. Unlike the Poisson distribution of a uniform random graph, which is sharply peaked around a well defined mean, power law distributions are highly skewed with long tails, corresponding to increased network heterogeneity. Fig. 4*b* implies that random graphs with power law degree distributions tend to be much less vulnerable to random shocks than uniform random graphs with the same z, a point observed elsewhere (19, 20) with respect to network connectivity. Although this distinction between threshold and network heterogeneity is slightly surprising (because both kinds of heterogeneity are related by the fractional threshold condition), it is understandable. Nodes that are vulnerable because of a low threshold can still be well connected to the network, making them ideal early adopters. But nodes that are vulnerable to small perturbations because they have very few neighbors are therefore also poorly connected; hence, they have difficulty propagating any influence.

Network heterogeneity has an additional, complicating effect. Although networks with highly skewed degree distributions are more stable overall, within the cascade window they display increased susceptibility with respect to initial shocks that explicitly target high-degree nodes (19), even though such nodes are unlikely to be vulnerable themselves. If instead of choosing an initial node at random, we deliberately target a node with degree k, then the probability of at least one of its neighbors being a part of the largest vulnerable cluster, and therefore the probability of triggering a cascade, is $P_k = 1 - (1 - S_v)^k$, where S_v is the strength of the vulnerable cluster—a prediction that is well fit by numerical data (not shown) for uniform random graphs. Near the boundaries of the cascade window, where S_v is small, $P_k \simeq kS_v$, implying that the ratio between the probability of a global cascade being triggered by the most connected node in the network (with $k = k_{max}$) and an average node (with $k = z$) is approximately k_{max}/z, which is a rough measure of the skewness of the degree distribution p_k. Networks with highly skewed p_k (such as uniform random graphs near the lower cascade boundary in Fig. 1, or those with power-law degree distributions) should therefore exhibit the property that their most connected nodes are disproportionately likely to trigger global cascades when chosen as initial sites. By contrast, networks in which p_k is sharply peaked, with rapidly decaying tails (such as near the upper boundary of Fig. 1) will not display this property. Numerical results for uniform random graphs (not shown) support this conclusion. Hence, the value of deliberately targeting highly connected initial nodes depends significantly on the global degree distribution, and therefore, in the case of uniform random graphs, whether the system is in its high-connectivity or low-connectivity regime.

Conclusions

Global cascades in social and economic systems, as well as cascading failures in engineered networks, display two striking qualitative features: they occur rarely, but by definition are large when they do. This general observation, however, presents an empirical mystery. Both power-law and bimodal distributions of cascades would satisfy the claim of infrequent, large events, but these distributions are otherwise quite different, and might require quite different explanations. Unfortunately a lack of empirical data detailing cascade size distributions prevents us from determining which distribution (if either) correctly describes which systems. Here we have motivated and analyzed a simple, binary-decision model that, under different conditions, exhibits both kinds of behaviors and thus sets up some testable predictions about cascades in real systems. When the network of interpersonal influences is sufficiently sparse, the propagation of cascades is limited by the global connectivity of the network; and when it is sufficiently dense, cascade propagation is limited by the stability of the individual nodes. In the first case, cascades exhibit a power-law distribution at the corresponding critical point, and the most highly connected nodes are critical in triggering cascades. In the second case, the distribution of cascades is bimodal, and nodes with average connectivity, by virtue of their greater frequency, are much more likely to serve as triggers. In the latter regime, the system displays a more dramatic kind of robust-yet-fragile quality than in the former, remaining almost completely stable throughout many shocks before exhibiting a sudden and giant cascade—a feature that would make global cascades exceptionally hard to anticipate. Finally, systemic heterogeneity has mixed effects on systemic stability. On the one hand, increased heterogeneity of individual thresholds appears to increase the likelihood of global cascades; but on the other hand, increased heterogeneity of vertex degree appears to reduce it. It is hoped that the introduction of this simple framework will stimulate theoretical and empirical efforts to analyze more realistic network models (incorporating social structure, for example) and obtain comprehensive data on the frequency, size, and time scales of global cascades in real networked systems.

This paper benefited from conversations with D. Callaway, A. Lo, M. Newman, and especially S. Strogatz. The research reported was conducted at the Massachusetts Institute of Technology Sloan School of Management under the sponsorship of A. Lo.

1. Gladwell, M. (2000) *The Tipping Point: How Little Things Can Make a Big Difference* (Little, Brown, New York).
2. Shiller, R. J. (1995) *Am. Econ. Rev.* **85,** 181–185.
3. Lohmann, S. (1994) *World Politics* **47,** 42–101.
4. Bikhchandani, S., Hirshleifer, D. & Welch, I. (1992) *J. Pol. Econ.* **100,** 992–1026.
5. Aguirre, B. E., Quarantelli, E. L. & Mendoza, J. L. (1988) *Am. Soc. Rev.* **53,** 569–584.
6. Schelling, T. C. (1973) *J. Conflict Resolution* **17,** 381–428.
7. Kosterev, D. N., Taylor, C. W. & Mittelstadt, W. A. (1999) *IEEE Trans. Power Systems* **14,** 967–979.
8. Sachtjen, M. L., Carreras, B. A. & Lynch, V. E. (2000) *Phys. Rev. E* **61,** 4877–4882.
9. Perrow, C. (1984) *Normal Accidents: Living with High-Risk Technologies* (Basic Books, New York).
10. Carlson, J. M. & Doyle, J. (1999) *Phys. Rev. E* **60,** 1412–1427.
11. De Vany, A. S. & Walls, W. D. (1996) *Econ. J.* **106,** 1493–1514.
12. Banerjee, A. V. (1992) *Quart. J. Econ.* **107,** 797–817.
13. Glance, N. S. & Huberman, B. A. (1993) *J. Math. Soc.* **17,** 281–302.
14. Arthur, W. B. (1989) *Econ. J.* **99,** 116–131.
15. Granovetter, M. (1978) *Am. J. Soc.* **83,** 1420–1443.
16. Glaeser, E. L., Sacerdote, B. & Schheinkman, J. A. (1996) *Quart. J. Econ.* **111,** 507–548.
17. Valente, T. W. (1995) *Network Models of the Diffusion of Innovations* (Hampton Press, Cresskill, NJ).
18. Arthur, W. B. & Lane, D. A. (1993) *Structural Change and Economic Dynamics* **4,** 81–103.
19. Albert, R., Jeong, H. & Barabasi, A. L. (2000) *Nature (London)* **406,** 378–382.
20. Callaway, D. S., Newman, M. E. J., Strogatz, S. H. & Watts, D. J. (2000) *Phys. Rev. Lett.* **85,** 5468–5471.
21. Keeling, M. J. (1999) *Proc. R. Soc. London B* **266,** 859–867.
22. Stauffer, D. & Aharony, A. (1991) *Introduction to Percolation Theory* (Taylor and Francis, London).
23. Newman, M. E. J. & Watts, D. J. (1999) *Phys. Rev. E* **60,** 7332–7342.
24. Sethna, J. P., Dahmen, K., Kartha, S., Krumhansl, J. A., Roberts, B. W. & Shore, J. D. (1993) *Phys. Rev. Lett.* **70,** 3347–3350.
25. Adler, J. (1991) *Physics A* **171,** 453–470.
26. Solomon, S., Weisbuch, G., de Arcangelis, L., Jan, N. & Stauffer, D. (2000) *Physica A* **277,** 239–247.
27. Watts, D. J. (1999) *Small Worlds: The Dynamics of Networks Between Order and Randomness* (Princeton Univ. Press, Princeton).
28. Shrager, J., Hogg, T. & Huberman, B. A. (1987) *Science* **236,** 1092–1094.
29. Bak, P., Tang, C. & Wiesenfeld, K. (1987) *Phys. Rev. Lett.* **59,** 381–384.
30. Strogatz, S. H. (2001) *Nature (London)* **410,** 268–276.
31. Bollobas, B. (1985) *Random Graphs* (Academic, London).
32. Newman, M. E. J., Strogatz, S. H. & Watts, D. J. (2001) *Phys. Rev. E* **64,** 02611.8.
33. Stanley, H. E. (1971) *Introduction to Phase Transitions and Critical Phenomena* (Oxford Univ. Press, Oxford).